Short Communication

In silico fusion of epsilon and beta toxin genes of *Clostridium perfringens* types D and B

Reza Pilehchian Langroudi[1], Khosrow Aghaiypour[*2], Mahdi Shamsara[1], Saied Ali Ghorashi[1]

[1]Department of Molecular Genetics, National Institute of Genetic Engineering and Biotechnology, Tehran, P.O. Box 14965/161, I.R. Iran [2]Department of Genomix and Genetic Engineering, Razi Vaccine and Serum Research Institute, P.O. Box 31975/148, Karaj, I.R. Iran

Abstract
Fusion protein technology represents the strategy to achieve rapid, efficient, and cost-effective protein expression. Epsilon and Beta toxins are the most potent Clostridial toxins and cause disease in animals. This study describes in silico fusion of *Clostridium perfringens* types D and B epsilon and beta toxin genes that was used for cloning in *E. coli*. The *etx* and *cpb* genes were retrieved from the GenBank and a fusion gene was designed to produce a chimeric fusion protein. Secondary and tertiary structures and specificities of fusion protein were determined by online software. Results showed that the designed fusion gene construction is suitable for chimeric fusion protein expression.

Keywords**:** *Clostridium perfringens*; epsilon toxin; Beta toxin; fusion; *In silico*

Clostridium perfringens is an important human and livestock pathogen. It is a Gram-positive, rod-shaped, anaerobic, spore-forming, heat-resistant bacterium of genus *Clostridium*. The spores persist in the environment, and often contaminate raw food materials. These bacteria are found in soil and mammalian feces (Rood *et al*., 1991). Individual strains produce subsets of toxins (Petit *et al*., 1999) and four of them iota (iA), alpha (*cpa*), beta (*cpb*) and epsilon toxins (*etx*) have been proposed to be used for classification of *Clostridium perfringens* into five isotypes A, B, C, D, and E (McDonel *et al*., 1986), where each type carries a different combination of the toxin genes. Alpha toxin (*cpa*) is found in *Clostridium perfringens* types A, B, C, D, and E, while the beta toxin (*cpb*) is found in types B and C. The epsilon toxin (*etx*) is found in types B, and D. The iota toxin (iA) is found only in type E (McClane *et al*., 2001).

Epsilon-toxin is a potent toxin and a member of a group of highly toxic Clostridial toxins. It is the third most potent Clostridial toxin after botulinum and tetanus neurotoxins (Payne *et al*., 1997). Epsilon toxin is normally produced as a relatively inactive prototoxin and is activated by proteolytic hydrolysis and conformational changes to cytotoxin, causing rapidly fatal enterotoxemia in livestock (Habeeb *et al*., 1973; Bhown *et al*., 1977). The main biological activity of epsilon toxin is the production of oedema in various organs. Epsilon toxin is known as one of the *Clostridium*/*Bacillus mosquitocidal* toxin ETX /MTX2 family members.

Beta toxin causes enterotoxaemia and necrotic entrities in lambs, piglets and calves (Hogh, 1976). In human, beta toxin is known to produce necrotic entrities (Granum, 1990). Beta-toxin gene sequencing has revealed similarity (17%-29%) to alpha-toxin, gamma toxin and Leucocidin from *Staphylococcus aurous* (Hunter *et al*., 1993). Leucocidin/ASH4 hemolysin, Domain and *staphylococcal* bi-component toxin, are beta channel forming cytolysins and the Clostridium

*Correspondence to: **Khosrow Aghaiypour, Ph.D.***
Tel: +98 261 4570038; Fax: +98 261 4552194
E-mail: khosrow@rvsri.ir

perfringens type B beta toxin is known as one of the members of these protein family.

A physical map of the genome of *Clostridium perfringens* strain CPN50, associated with human disease has been shown to possess a single circular chromosome of about 3.58 Mb by PFGE. More than 100 restriction sites and 24 genetic loci have been located on the genome (Canard *et al*., 1989). The complete 3031430 bp genome sequence of *Clostridium perfringens* strain 13 comprises of 2660 protein coding regions and reported 10 rRNA genes, that shows pronounced low overall G+C content (28.6%) (Shimizu *et al*., 2002).

Sequences retrieval and chimeric fusion gene and fusion protein construction: The complete nucleotide sequences of *etx* and *cpb* were retrieved from the GenBank. The first data used was GenBank accession number HQ179546 (*Clostridium perfringens* epsilon toxin gene partial cds), as the representative sequence of *etx* gene. Analysis of this nucleotide sequence showed that it codes for a protein containing of 328 amino acids. A fragment of 32 amino acids at the N-terminal is a signal peptide and the next 296 amino acids fragment is an unactivated mature peptide or prototoxin. The same steps were performed on GenBank accession numbers HQ179547, and HQ424445 (*Clostridium perfringens* beta toxin gene partial cds), as the representative sequences of *cpb* gene. Analysis of nucleotide sequence showed that it codes for a protein containing of 336 amino acids. A fragment of 27 amino acids at the N-terminal was found to be a signal peptide and the remaining 309 amino acid fragment was a mature peptide.

The nucleotide sequences were used to design a chimeric fusion gene which is constructed by *etxD* and *cpbB* linked together via the linker fragment AEAAAKEAAAKA. *Nde*I restriction site and it's flanking region at the 3′ end of *etx* gene and *Xho*I restriction site and it's flanking region for the 5′ end of *cpb* gene also was designed. The designed fusion gene length is 1947 bp, which within 1 to 984 is *etx*. Base pairs 985 to 1020 (36 bp) are linker sequence, which is optimized for *E. coli* (GCGGAAGCGGCGGCGAAAGAAGCGGCGGCGAAAGCG). Base pairs 1021 to 1947 (927 bp) are *cpb*. Figure 1 shows both schematic view of designed fusion gene and fusion chimeric protein construction.

Online analysis of fusion protein: Patterns and profiles of the fusion chimeric protein were predicted by Interpro online software (http://www.ebi.ac.uk/Tools/InterProScan/) and revealed that the whole construction is made up of *Clostridium perfringens* epsilon toxin ETX/*Bacillus mosquitocidal* toxin MTX2, family member which is combined with Leucocidin/ASH4 hemolysin domain and also *Staphylococcal* bi-component toxin family members. Figure 2 shows Interpro scan for fusion chimeric protein.

Secondary structure was predicted using v.3.0 of Gail Hutchinson's PROMOTIF program (http://www.rubic.rdg.ac.uk/~gail/#Software). Three sets of analysis were prepared using epsilon toxin peptide (328 amino acids), beta toxin peptide (309 amino acids) and the fusion chimeric peptide (649 amino acids). The result showed that total residues are 649 which is made up of 9 sheets, 9 beta hairpins, 1 psi loop, 7 beta bulges, 34 strands, 8 helices, 2 helix-helix inter-

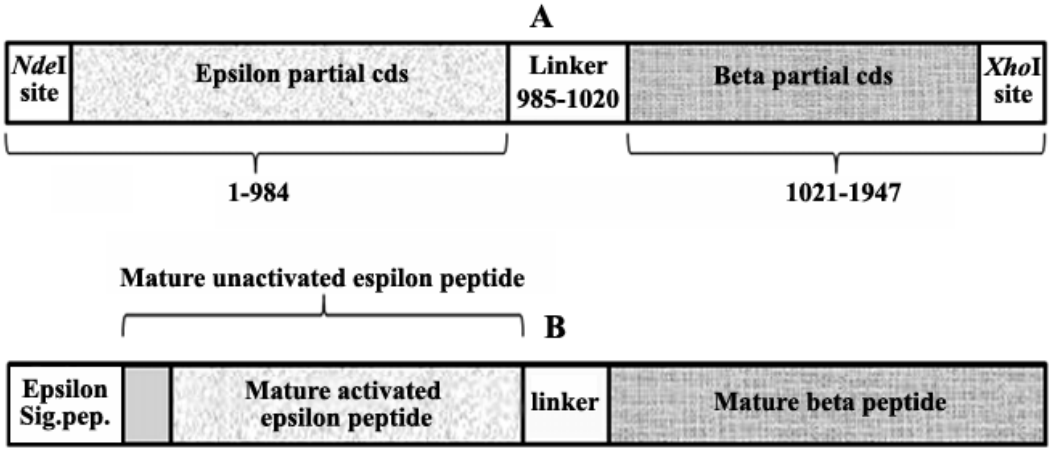

Figure 1. A: schematic view of designed fusion gene, B: schematic view of fusion chimeric protein construction.

acts, 57 beta turns and 43 gamma turns. Figure 3 shows secondary structure of fusion protein construction. GOR was used for the location of linker fragment (Garnier *et al*., 1996). The results showed that there is a helix pick located between positions 328-340 that corresponded to the linker fragment Figure 4. SignalP prediction was used for locating the signal peptide cleavage sites based on neural networks (NN) trained on Gram-positive bacteria. Accordingly, most likely cleavage site is located between positions 32-33. Figure 5 shows predicted signal peptide cleavage sites. HMMTOP was carried out for prediction of the transmembrane helices location in the fusion protein, which determined that fusion protein N-terminus is IN and there is transmem-

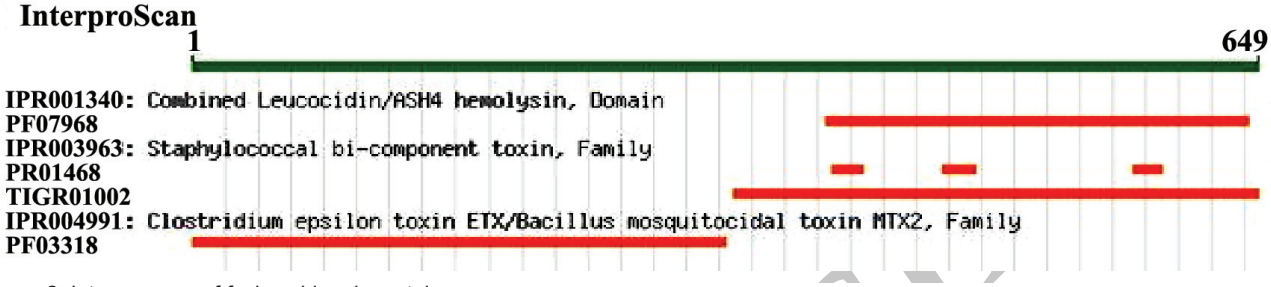

Figure 2. Interpro scan of fusion chimeric protein.

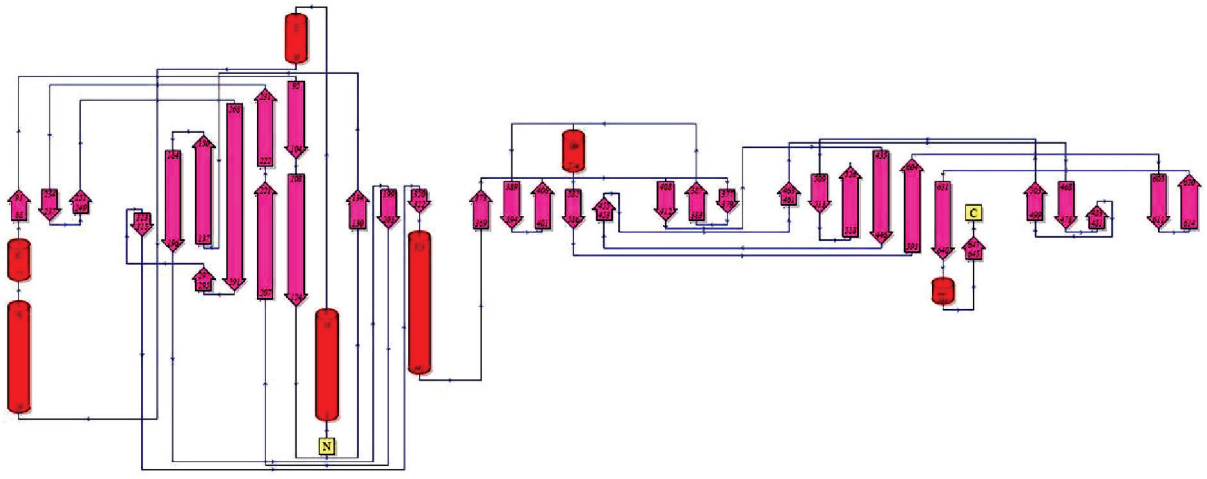

Figure 3. Secondary structure of the fusion protein construction.

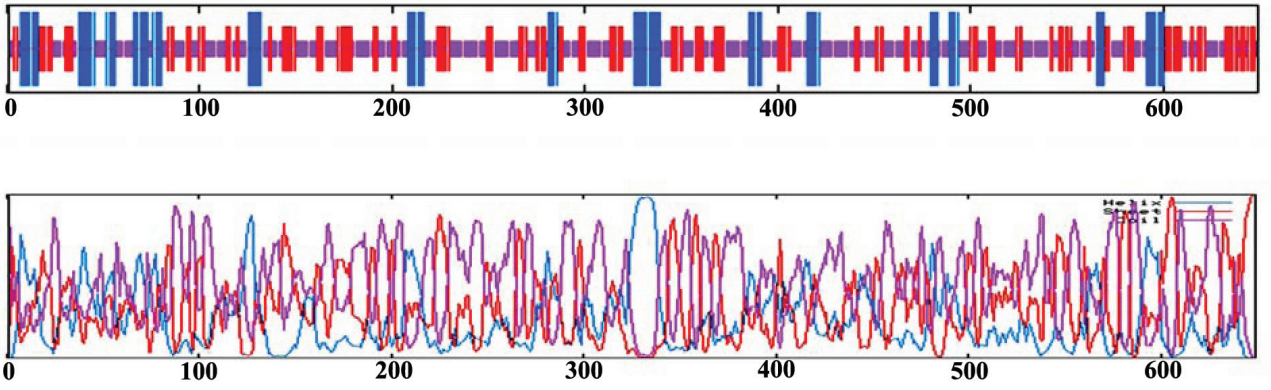

Figure 4. Secondary structure analysis of fusion protein construction by GOR. The linker fragment located at amino acids number 329 to 340. Amino acids 1 to 328 presents the epsilon toxin and amino acids 341 to 649 are belong beta toxin.

brane helices which is located between amino acids number 8 and 25.

Tertiary structure prediction of the fusion protein: Tertiary structure of the fusion protein was predicted by uploading several models to the I-TASSER server (Zhang *et al*., 2008). Uploaded models included complete epsilon toxin (328 amino acids), epsilon prototoxin peptide (296 amino acids), epsilon activated toxin (283 amino acids), complete beta toxin (336 amino acids) and beta mature toxin (309 amino acids). Three chimeric fusion construction models were also uploaded including model 1 (complete epsilon toxin + linker 1 + beta mature toxin, containing 649 amino acids), model 2 (complete epsilon toxin + linker 2 + beta mature toxin, containing 654 amino acids) and model 3 (complete epsilon toxin + linker 3 + beta mature toxin, containing 652 amino acids). Linker 1, 2 and 3 respectively were AEAAAKEAAAKA; AEAAAKEAAAKEAAAKA; GGGGSGGGGSGGGGS (Ryoichi *et al*., 2001). Tertiary predication result of the fusion protein construction showed a protein with two domains linked together with a small linker.

Circumstantial evidence supports the hypothesis that the assembly of relatively small domains into functional proteins is an important factor in evolution. Recombinant DNA technology has allowed *in vitro* fusions of genes or gene fragments in a simple and predictable manner. Fusion of *Clostridium perfringens* type D and B epsilon and beta toxin genes and it's cloning in *E. coli* was reported previously (Pilehchian Langroudi *et al*., 2011) .In the present study, the bioinformatics approach that was used for in silico fusion of epsilon and beta toxin genes of Clostridium perfringens types D and B is described. According to the latest information, this is the first design of epsilon-beta fusion gene for the development of a recombinant fusion protein. These findings indicated that the designed fusion protein could be used for vaccine production to prevent toxicosis of Clostridium perfringens epsilon and beta toxin. *Clostridium perfringens* is a spore forming bacterium and widely occurring pathogen. It is several decades that large-scale vaccine against epsilon and beta toxins is manufactured in Iran (Ardehali and Darakhshan, 1974). Large-scale production of vaccines from virulent strains requires stringent safety conditions and costly detoxification and control steps. Therefore, it would be beneficial to produce these toxins in a safe production host and in an immunogenic form. Bacterial expression systems are the most convenient to use for this purpose. Recombinant protein expression levels is generally high in bacteria. Bacterial expression systems are suitable to use when no post-translational modifications of the protein are required. Several bacterial systems are available, however, *E.coli* is widely used and also the best characterized (Dertzbaugh, 1998). In the present study, the complete epsilon toxin gene sequence including it's signal peptide for the proper secretion of the fusion protein, and the beta toxin gene sequence

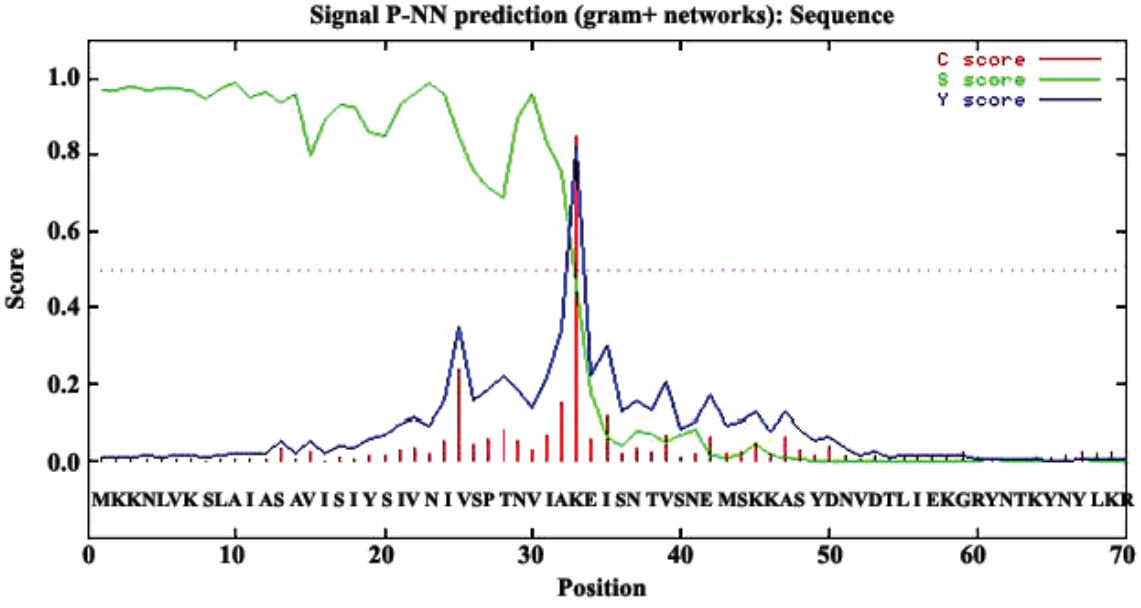

Figure 5. Prediction of signal peptide cleavage sites.

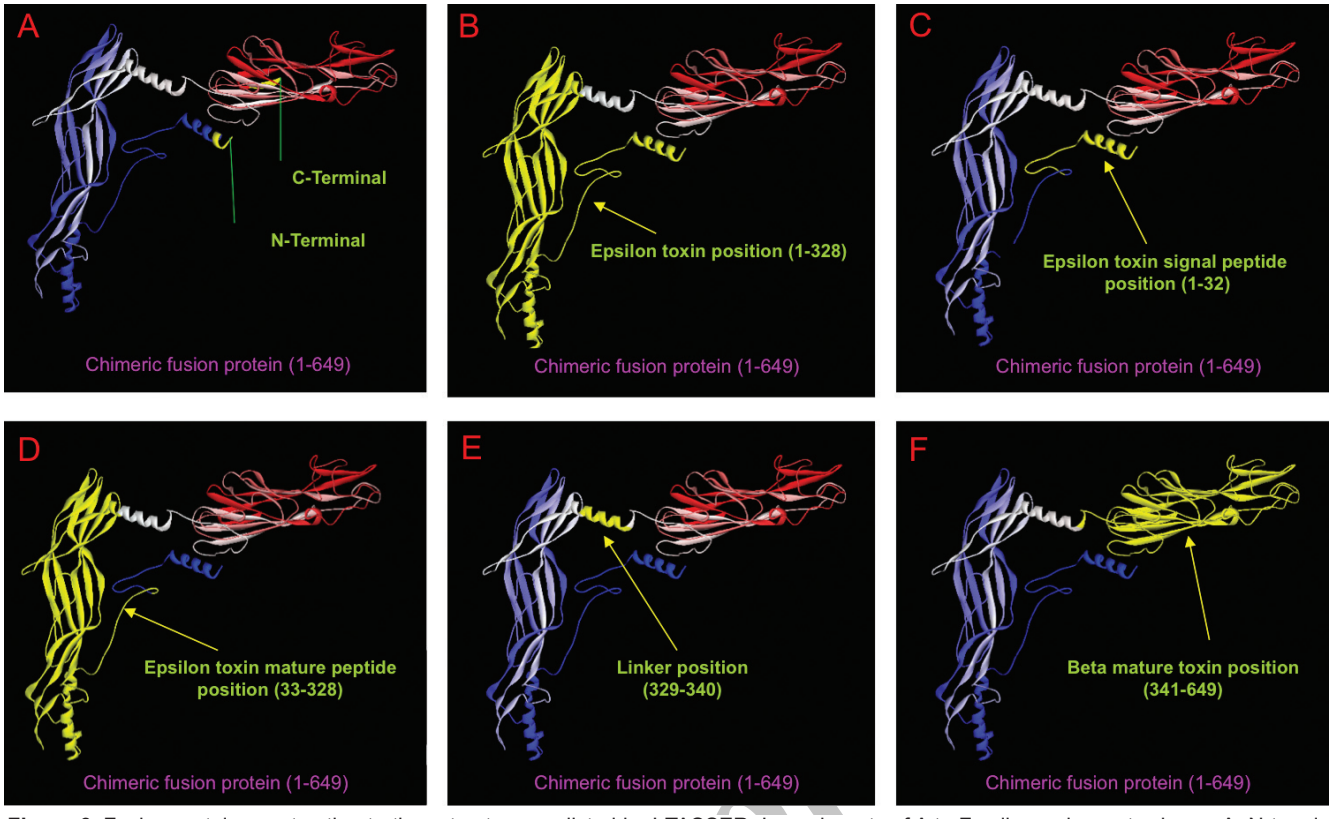

Figure 6. Fusion protein construction tertiary structure predicted by I-TASSER. In each parts of A to F yellow color parts shows: A, N-terminal of the epsilon toxin and C-terminal of beta toxin; B, complete epsilon prototoxin (328 amino acids); C, epsilon toxin signal peptide (32 amino acids); D, epsilon toxin mature peptide (295 amino acids); E, linker (12 amino acids); F, beta mature toxin (308 amino acids).

without it's signal peptide was used. Two repeats of "EAAAK" as a linker was designed. The InterProScan search identifies sequence motifs from several databases including PROSITE, PRINTS, Pfam-A, TIGRFAM, PROFILES and PRODOM. According to these data bases, the whole structure of the fusion protein construction is made up of *Clostridium* epsilon toxin ETX/MTX2 family member which is combined with Leucocidin/ASH4 hemolysin domain.

In 1992 Hunter *et al*. showed that the nucleotide and derived amino acid sequences of the epsilon toxin gene are located between the start codon at the base 188 and the stop codon at the base 1174 (986 bp). These data demonstrated that the amino acid sequence from the start Met to Lys (base 284) about 32 amino acid, contained residues characteristic of a signal peptide derived from the *etx* gene suggesting that Lys may be the first residue of the mature, inactivated toxin. In another study in 1993, Hunter *et al*. showed that the beta toxin, which is produced by *Clostridium perfringens* types B and C, is synthesized as a 336 amino acid protein and the first 27 residues of which constituted a signal peptide.

For confirmation of this finding, secondary structures of epsilon and beta toxins and the fusion chimeric protein were predicted by PROMOTIF (Hutchinson *et al*., 1996) online program software. Our data demonstrated similar characteristics of each of the epsilon and beta fragments comprising of epsilon-beta fusion protein constructions. The data, which is showed in Figure 6, confirms that amino acid Lys number 33 is the first residue of the epsilon mature, inactivated toxin.

Bhown and Habeeb reported that the trypsin cleavage site in epsilon-prototoxin is located between Lys-14 and Ala-15. Hunter *et al*., data derived from the epsilon toxin gene indicated that the residue preceding Lys-14 was also Lys. Since trypsin cleaves proteins on the C-terminal side of basic amino acids, this suggested that there were two possible sites for the cleavage which could explain this finding that the N-terminal residue of epsilon-toxin was Lys-14, not Ala-15. So from Lys (base 284) up to Lys (base 323) about 13 amino acids would be removed after cleavage.

Based on the present study results, for the first time the designed chimeric gene was constructed and ligated into pJET1.2blunt cloning vector and *E. coli* strain TOP10 was transformed by this DNA insert. (Pilehchian Lnagroudi *et al*., 2011). The sequence of cloned fusion gene was submitted to GenBank under accession number JF833085. As found in NCBI protein ID AEH25946.1, translation of the fusion gene in the cytoplasm of transformed host cell would produce a 649 amino acids chimeric epsilon-beta fusion protein (synthetic construct), which consists of epsilon toxin signal peptide (32 amino acids), mature inactivated peptide (296 amino acids) the small hydrophobic linker (12 amino acids) and beta toxin mature peptide (309 amino acids). The *Nde*I and *Xho*I restriction sites and their flanking regions at the 3' end of epsilon and 5' end of beta toxin genes, which respectively are necessary for insertion of fusion construction into PET system expression vector, are present in the synthetic construct.

The InterPro scan search findings were confirmed by the results of tertiary structure prediction of the fusion protein construction using I-TASSER online program. Results demonstrated a protein with two main domains linked together with a small linker. The secondary structure prediction results were also confirmed by tertiary structures prediction results of complete epsilon toxin, epsilon prototoxin, epsilon activated toxin, complete beta toxin, beta mature toxin and three models of fusion proteins. Results showed that secondary structures specificities of both of epsilon and beta are present in tertiary structure of fusion genes, but the best C-score (-3.075), 0.37±0.12 (TM-score), 15.7±3.3Å (RMSD) belonged to model 1 which has linker-1.

Tertiary structural of different parts and domains of fusion chimeric protein construction, which was predicted by I-TASSER, are shown in Figure 6. The hydrophobic linker is introduced between two domains which can provide suitable flexibility and separation. After expression, these constructions can provide suitable accumulation of fusion protein in the periplasmic space of *E. coli*. This will happen because there is a signal peptide at the N terminal of the epsilon toxin of chimeric protein, which will allow it to cross the cytoplasmic membrane. As the expressed fusion protein construction will be secret to the culture media, thus inclusion body formation would not occur. This statement is further supported by the fact that native epsilon toxin gene encoded a sequence, which was structurally similar to a signal peptide, and it is known that the *Clostridium perfringens* native toxins are secreted to the culture media.

Protein models, which predicted by I-TASSER are generated by combining the methods of threading, ab initio modeling and structural refinement (Roy *et al*., 2010; Zhang 2009; Zhang 2008). The procedure is fully automated and the final model is generated as PDB format. We uploaded this model to the profunc server which had been developed to help identify the likely biochemical function of a protein from it's three-dimensional structure. The Profunc server uses both sequence and structure-based methods in attempt to provide clues for the protein's likely or possible function (Laskowski *et al*., 2005 a,b). Often, where one method fails to provide any functional insight, another may be more helpful. Using a prediction program can only get close to the final properties and specificities of a desired designed protein and it does not exactly reflect the absolute and definite properties and specificities of that protein. Besides, in silico study of a functional protein folding and construction is the most important and the best way to understand it's structure and functions. Results of the present study clearly show that the designed epsilon-beta fusion gene construction could be produced and expressed in a suitable host cell.

Acknowledgments

Financial and scientific supports provided by National Institute of Genetic Engineering and Biotechnology and Razi Vaccine and Serum Research Institute is acknowledged.

References

Ardehali M, Darakhshan H (1974). Production and standardization of *Clostridium* perfringens polyvalent vaccine in Iran. *Develop Biol Standard*. 32: 31-34.

Bordoli RK, Kopp LJ, Schwede T (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modeling. *Bioinformatics* 22: 195-201.

Brown AS, Habeeb AFSA (1977). Structural studies on epsilon toxin of *Clostridium perfringens* type D. Localisation of the site of tryptic scission necessary for activation of epsilon toxin. *Biochem Biophysl Res Commu*. 73: 889-896.

Canard B, COLE ST (1989). Genome organization of the anaerobic pathogen; *Clostridium perfringens*. *PNAS*. 86: 6676-6680.

Dertzbaugh MT (1998). Genetically engineered vaccines: an overview. *Plasmid* 39: 100-113.

Garnier JF, Gibrat B, Robson GOR (1996). Secondary structure prediction method version IV, *Methods in Enzymology*, RF Doolittle Ed. 266: 540-553.

Habeeb AFSA, Lee CL, Atassi MZ (1973). Conformational studies on modified proteins and peptides VII Conformation of

epsilon prototoxin and epsilon toxin from *Clostridium perfringens*. Conformational changes associated with toxicity. *Biochemica et Biophysica Acta*. 322: 245-250.

Hogh P (1976). Experimental studies on serum treatment and vaccination against *Clostridium perfringens* type C infection in piglets. *Dev Bio Stand*. 32: 6-76.

Granum PE (1990). *Clostridium perfringens* toxins involved in food poisoning. *Int.J.Food Microbial.* 10: 101-111.

Hunter SEC, Clarck IN, Kelly DC, Titball RW (1992). Cloning and Nucleotide Sequencing of the *Clostridium perfringens* Epsilon-Toxin Gene and Its Expression in *Escherichia coli*. *Infect Immun*. 60: 102-110.

Hunter SE, Brown E, Oyston PCF, Sakurai J, Titball RW (1993). Molecular genetic analysis of beta-toxin of *Clostridium perfringens* reveals sequence homology with alpha-toxin, gamma-toxin, and leukocidin of Staphylococcus aureus. *Infect Immun*. 61: 3958-3965.

Hutchinson EG, Thornton JM (1996)."PROMOTIF- A program to identify structural motifs in proteins" *Protein Sci.* 5: 212-220.

Laskowski RA, Watson JD, Thornton JM (2005 a). ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res*. 33: W89-W93.

Laskowski RA, Watson JD, Thornton JM (2005 b). Protein function prediction using local 3D templates. *J Mol Biol*. 351: 614-626.

McClane BA (2001). *Clostridium perfringens*. Doyle MP, editor. Food microbiology: *fundamentals and frontiers*. Washington, DC: ASM Press. PP. 351-372.

McDonel JL (1986). Toxins of *Clostridium perfringens* types A, B, C, D, and E, Dorner F, Drew H, editors. *Pharmacology of bacterial toxins*. Oxford: Pergamon Press. PP. 447-517.

Nagahama M, Sakurai J (1991). Distribution of labeled *Clostridium perfringens* epsilon toxin in mice. *Toxicon* 29: 211-217.

Nagahama M, Sakurai J (1992). High-affinity binding of *Clostridium perfringens* epsilon-toxin to rat brain. *Infect Immun*. 60: 1237-1240.

Oakley CL (1943). Toxins of Cl. welchii; a critical review. *Bull Hyg*. 18: 781-806.

Oakley CL, Warrack H (1953). Routine typing of *Clostridium welchii*. *J Hyg*. 51: 102-107.

Payne DW, Oysten P (1997). The *Clostridium perfringens* epsilon toxin. *The Clostridia*: *Molecular biology and pathogenesis*. Academic Press Limited: London. PP. 439-447.

Petit L, Gibert M, Popoff MR (1999). *Clostridium perfringens*: toxinotype and genotype. *Trends Microbiol*. 7: 104-110.

Pilehchian Langroudi R, Aghaei Pour K, Shamsara M, Jabbari AR, Habibi GR, Goudarzi H, Ghorashi SA (2011). Fusion of *Clostridium perfringens* type D and B epsilon and beta toxin genes and it's cloning in *E. coli*. *Archives of Razi Institute*. 66: 1-10.

Rood JI, Cole ST (1991). Molecular genetics and pathogenesis of *Clostridium perfringens*. *Microbio Rev*. 55: 621-648.

Roy A, Kucukural A, Zhang Y (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nature Protocols*. 5: 725-738.

Ryoichi A, Hiroshi U, Astushi K, Noriho K, Teruyuki N (2001). Protein Design of linkers which effectively separate domains of bifunctional fusion proteins. *Engineering*14: 529-532.

Shimizu T, Ohtani K, Hirakawa H, Ohshima K, Yamashita A, Shiba T, Ogasawara N, Hattori, M,Kuhara S, Hayashi H (2002). Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *PNAS*. 99: 996-1001.

Uhlén M (1989). Dual affinity fusion approach and its use to express recombinant human insulin-like growth factor II. *PNAS.* 86: 4367-4371.

Zhang Y (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 9: 1-8.

Zhang Y (2009). I-TASSER: Fully automated protein structure prediction in CASP8. *Proteins* S9: 100-113.

Zdobnov EM, and Apweiler R (2001). InterProScan an integration platform for the signature-recognition methods in InterPro, *Bioinformatics* 17: 847-848.